Desktop GIS for Environmental Sensitivity Index Mapping

Background

In 1996, the first CD of digital environmental sensitivity index (ESI) data for Geographic Information Systems (GIS) was published by NOAA/HAZMAT. Geographic data were provided in ARC export format together with the data base files that had been utilized for data compilation and map generation. It soon became apparent that this format was only suitable for the high-end users, i.e. those who would be responsible for data base maintenance and updates. Significant manipulation and simplification of the data base was required by the users who wished to create and display custom map products, and those who wished to perform simple query and analysis of the data. This manipulation required users have access to sophisticated data base systems and that they become intimately familiar with a data base structure that is anything but intuitive. The consequence was duplication of effort by those who purchased the CD, erroneous interpretation of the data, and many who simply abandoned the idea of using ESI data in their GIS. The inability to use these valuable data was frustrating to both the users and NOAA. In an attempt to remedy the situation, HAZMAT has invested significant effort in developing a data set that would meet the needs of the growing number of desktop GIS users. This effort, which also involves additional quality checks for data consistency and usability, is discussed in the following pages. Suggestions and comments on the simplified data structure as well as the formats used to distribute the geographic files are encouraged. Please contact Jill Petersen at j_petersen@hazmat.noaa.gov.

Geographic Data

Environmental Sensitivity data can be divided into three main categories: base map, biology and human use. The base map layers include **hydro**, **index**, and **esi**. Biology includes some or all of the following layers: **birds**, **nests**, **fish**, **t_mammal**, **m_mammal**, **reptiles**, **invert** and **habitats**. The human use layers are **mgt** and **socecon**. For details on

content of these layers see the *Environmental Sensitivity Index Guidelines*. For details on how the geographic data is distributed and how it links to the data base files, reference this document.

All geographic data are provided in geographic coordinates with the horizontal datum being NAD27. When data are provided in ARC export format, the files are exported without compression. These parameters were selected as, in addition to being suitable for the large number of ARC/INFO and ArcView users, there are also translators available for converting uncompressed ARC files to formats suitable for import into other mapping programs.

All geographic files are distributed in four different formats. The first of these is termed **source** data. These data are double precision ARC export files. These files most closely resemble the data as described in the *Environmental Sensitivity Index Guidelines*. Though they undergo the same quality control checks as all files, and in the case of the biology layers some minor modifications are made (see **Biology** heading), they are distributed in the same precision and have the resolution of the data used to generate the hard copy maps. Users who will be responsible for maintaining and updating ESI data should use these data in a GIS system that supports double precision data.

The second format used for distribution is termed **desktop**. These files are also in ARC export format, but are exported as single precision files. For users of mapping programs that only support single precision data (such as ArcView 3), these files should be used rather then letting the mapping program do the precision conversion. This is because polygons may collapse when the precision is reduced. Before the desktop files are exported, polygons that collapsed are checked and corrected to assure all polygons are labeled properly. Some of the very large polygons in the hydro layer (and occasionally the esi layer) exceed the 5000 arc limit imposed by ArcView 3, PC ARC/INFO and some other mapping programs. These polygons are subdivided and appropriately labeled to avoid this problem. When this is done, the added arcs' line attribute is assigned an 'I', to indicate it is an unnatural boundary the same as the index lines representing quad or study boundaries.

Users with a system that supports double precision data may still benefit

from using some of the desktop files in lieu of the source files. For example, they may choose to use the full resolution data for the detailed hydro and esi layers, but wish to use the desktop files for biology and human use data. Choosing this method may increase drawing time significantly with minimal compromise in data quality.

The third distribution format is **MOSS** formatted files. MOSS is a simple ascii format that supports topology. MOSS is supported by various mapping programs including the public domain program, MOSS, and the commercial program CAMRIS. A description of the data format is attached at the end of this document for users who may want to write their own translators to bring ESI data into their system. The content of the MOSS files is equivalent to the desktop files, i.e. they are single precision files which meet the maximum 5000 arc limit.

The final distribution format is **MARPLOT**. Geographic data are provided in MARPLOT format bundled with the desktop data files already imported into its supporting data base, OMO. For a brief description of the MARPLOT program, see attached documentation. There is also a readme file provided on the CD that explains how to organize the MARPLOT files on your hard drive. MARPLOT will meet the needs of users who wish to do simple query and analysis of the ESI data, but who do not have an established GIS/desktop mapping program in place. It may also be useful for other users to check data brought into their own mapping programs for proper linkage to the data base as well as general presentation considerations such as standard color and hatch patterns.

Data Attribute Files

In an attempt to support a wide range of users, data attribute files are distributed in both comma delineated text format and as uncompressed ARC export files. They are provided in a simplified "desktop" format, as well as a collection of data files for use in a relational data base closely resembling the data files described in the *Environmental Sensitivity Index Guidelines*. With the exception of those responsible for updates to the ESI data, the simplified data format will meet the needs of most users. It should be recognized, however, that the desktop format is not normalized and changes made to the desktop files cannot revert to the relational structure. The desktop format is most suitable for users wanting to view, query and analyze the ESI data. On the desktop ESI CDs, all data files are placed in a directory called dbdir. In this directory there are two sub directories, desktop and relfiles, each of which has an additional two sub directories, text and export. The contents of the sub directories is identical information in varying formats.

Biology

There are two primary **desktop** data files that link to the biology layers. The first, *bio_lut*, will not be needed if the mapping program used does not require that each polygon have a unique id and hence can support a many to many link to the data base. The other, *biofile* contains a summary of all the attribute data as well as links to two additional files *sources* and *breed*. These additional files are provided for users who are using a data base that allows multiple links. The *sources* file provides information about who collected the data, vintage of the data and publication information. This information is duplicated in the metadata file (also provided on the CD), though not in a polygon specific manner. The *breed* file provides life activity information in a searchable format. The utility of this file would be recognized when a user wants to select only the fish polygons where spawning is occurring in February. Though not in a searchable format, this breed activity is summarized in the *biofile.*

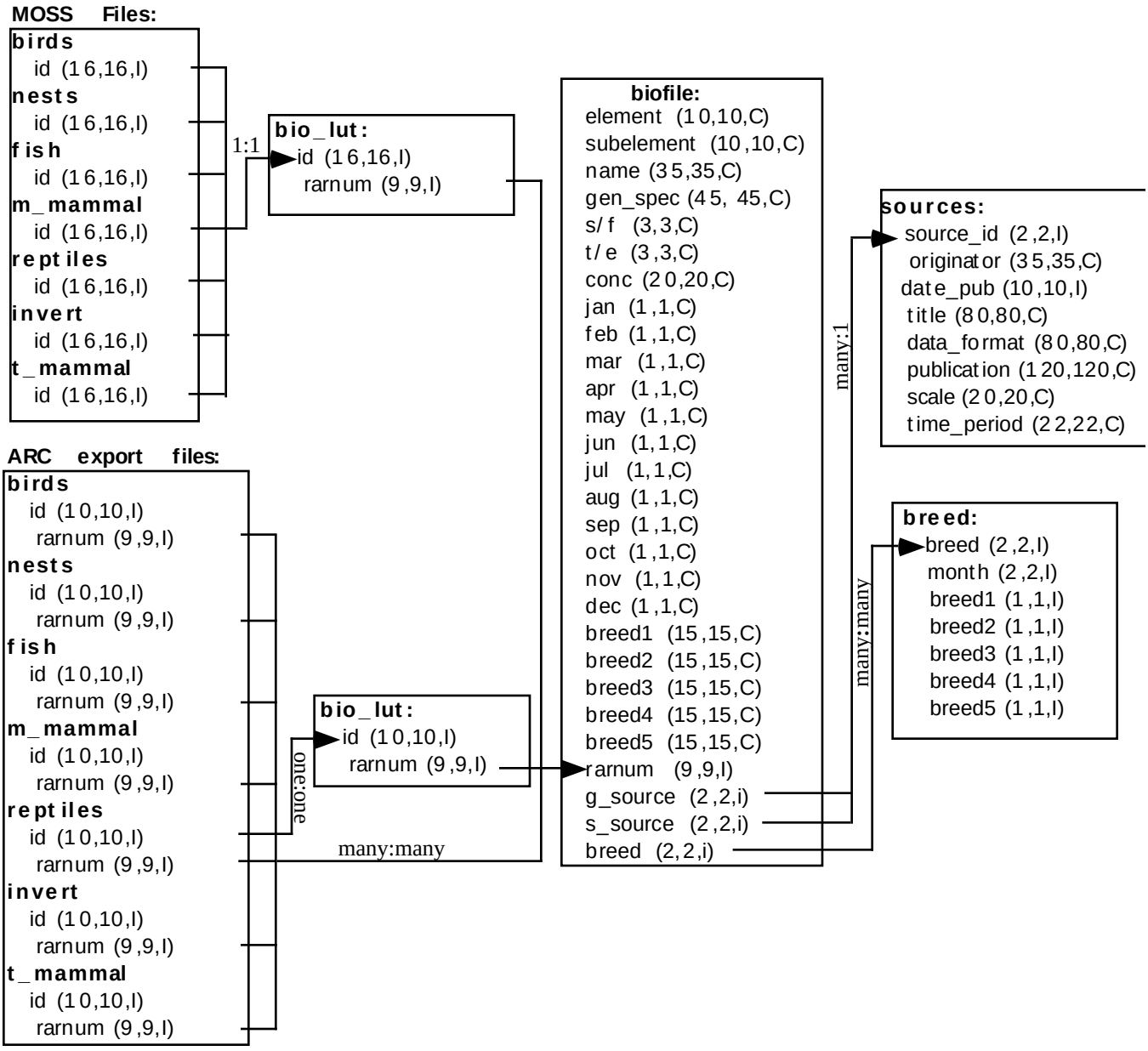


Figure 1: The relationship of the ARC and MOSS files to the biofile (The ARC files refer to either the double precision source files or the single precision "desktop" files.)

Figure 1 shows the linkage from the MOSS and ARC files to the attribute data. The ESRI products ARC/INFO and ArcView 3 support many to many links so the *bio_lut* may be bypassed. However the id is retained in these files for those using a translator to bring these data into mapping programs that may support only a one to one link from the geography. MOSS files support only one attribute per layer so the id was exported to assure

that all users needs would be met.

Note that the table *bio_lut* appears twice in figure 1, with varying definitions for id. Both the id items are generated by concatenating a unique polygon identifier, the layer identifier and the atlas number. The MOSS files, which are also used for MARPLOT import, additionally concatenate the rarnum with the id. Because ARC does not support integers wider than 10, this value is not incorporated into the ARC export id. It is important that MOSS users use the proper bio_lut file. It resides in the directory geodir/moss and is called bio_lut.txt.

The rarnum in both the desktop and the source data files differs slightly from the rarnum described in the *Environmental Sensitivity Index Guidelines*. Digital files used to generate the hard copy maps are reprocessed to generate new rarnums that are specific to an element rather then passing through multiple layers. This supports the theory that digital files will be queried by layer. The element specific rarnum eliminates the confusion that would arise when reptiles appear as being present in a polygon when a user is querying the bird layer. Once the new rars are generated, adjacent polygons that have the same species present (as well as the other criteria defining an rar) within a particular element, are joined to become one large polygon. Additionally rarnums in the files distributed by NOAA will be unique across atlases. This allows the merging of multiple atlases and attribute files without the need to revert back to the always unique id. The rarnum is a 16 digit number composed of the atlas number, layer/element number and the unique group number.

The first 19 items in *biofile* (items element through dec) summarize information described in the *Environmental Sensitivity Index Guidelines* from the tables *biores*, *seasonal* and *species*. The items *breed1 - breed5* present a textual summary of the life activity data found in the *breed* table described in this manual. This is presented for easy reference as well as to support the needs of users who are unable to link to the newly defined *breed* table provided. Definitions of each activity by element are described in the *Guidelines* as well as a description of when there will be five breed variables available vs. four. The last three items in the *biofile* are the links to and from the *biofile*. It should be noted that the rarnum key is a many to many link. There may be many polygons that have the same rarnum and there may be many records for each rarnum. At a

minimum, there will be one *biofile* record for each species present in an rarnum grouping. If a polygon represents a "hole" in the cover, for example, when there is a land polygon contained inside a fish polygon, it will link to one record where the name field states "hole".

Likewise, the link to the *breed* table is many to many. Many species within the same element may share the same monthly life activities. Each breed variable will have twelve records in the *breed* table, one for each month, regardless of whether there is any special life activity present that month. Note that this is different from the breed table discussed in the *Guidelines* where like activity schedules are duplicated for each species regardless of uniqueness and where records are not included if no special life activity is present for a month. The approach taken here disallows redundancy and thus reduces the size of the *breed* table significantly. It also assures all *biofile* records link to a record in the *breed* table.

As previously mentioned, the *sources* file provides information about who collected the data, vintage of the data and publication information. This table is exactly the same as the *sources* table described in the ESI guidelines. For additional information on the contents of the data fields, please reference this document. There is a many to one link between the *biofile* and the *sources* file.

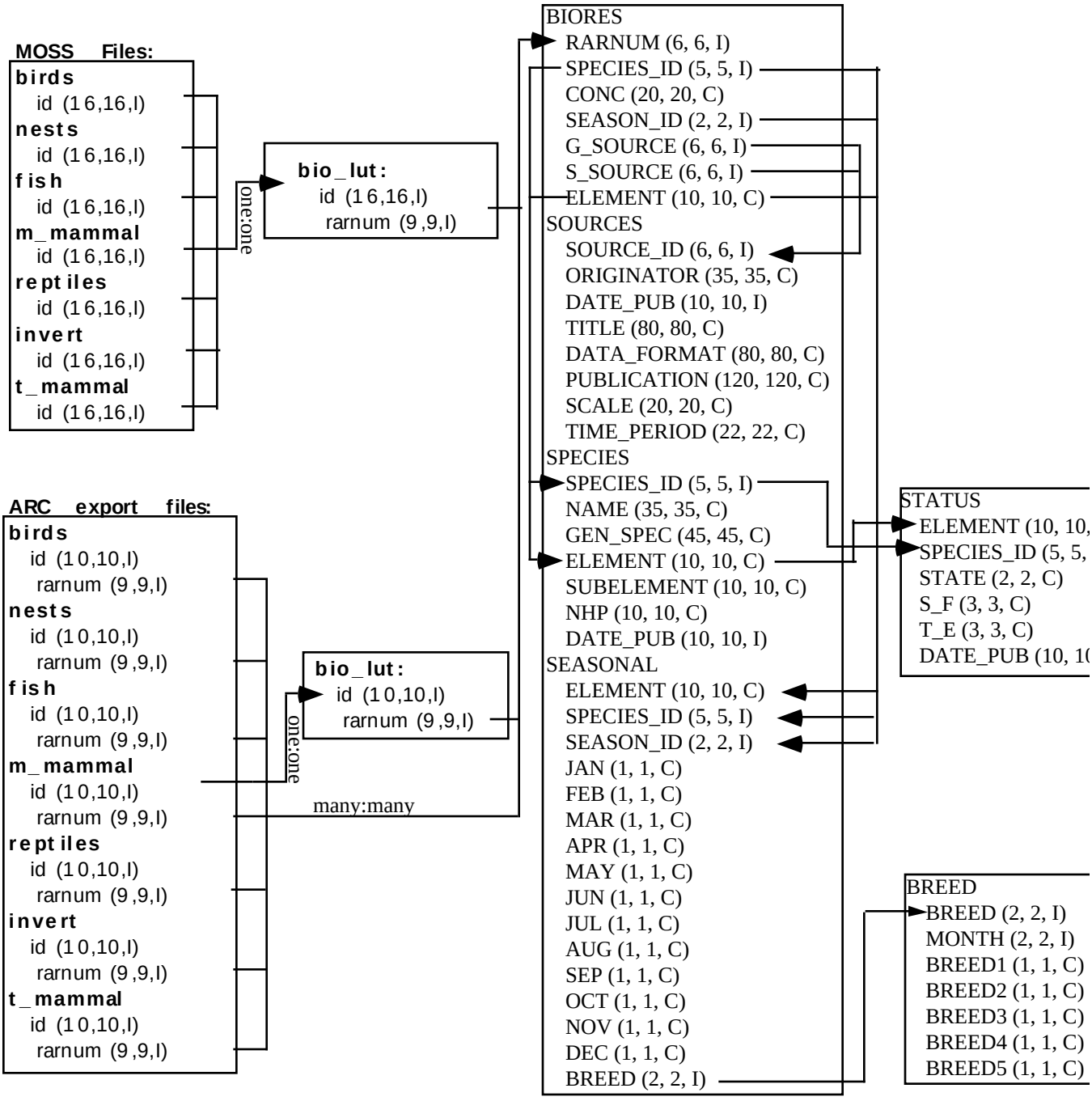


Figure 2: Navigating through the relational data structure (The ARC files refer to either the double precision source files or the single precision "desktop" files.)

The **relational files** found in the **relfiles** directory closely resemble the relational structure described in the ESI Guidelines. Figure 2 diagrams the link to the data base using either the id or the rarnum as well as the

multiple links between the data files themselves. For the files *biores*, *sources*, *species*, *seasonal*, and *status* please reference the ESI Guidelines for a description of content. Note that the lookup table here is *bio_lut*. Unlike the lookup tables described in the guidelines, this file combines all of the id/rarnum lookups regardless of whether they refer to point, polygon or arc data. The *bio_lut* file found in the relfiles directory is exactly the same as the *bio_lut* file used in the desktop version described above. Also, as described above, MOSS users should use the *bio_lut.txt* file found in the geodir/moss directory. The rarnum is unique across atlases as well as elements, as described above.

The *breed* table used with the relational files is the same as the *breed* table used with the desktop files. Please see the description of this table above as well as a discussion of how this table differs from the *breed* table described in the *Environmental Sensitivity Index Guidelines*.

Human Use

Attributes relating to the human use layers of the ESI are found in the file *soc_dat.* The link from the geography to this data table may be through either the id or humnum item. As described in the ESI Guidelines, the humnum is the key id to a record that may describe multiple polygons or points found in the atlas. This number has been modified from that described in the Guidelines to be unique across atlases, similar to the type of modification discussed for the rarnum. Also like the biology files, there are two versions of the lookup table, *soc_lut*. The lookup table for the MOSS files is located in the *geodir/moss* directory and is called *soc_lut.txt.* The desktop and relfile lookup table aer in the dbdir directory. For a description on the contents of the *soc_dat* table reference the Guidelines.

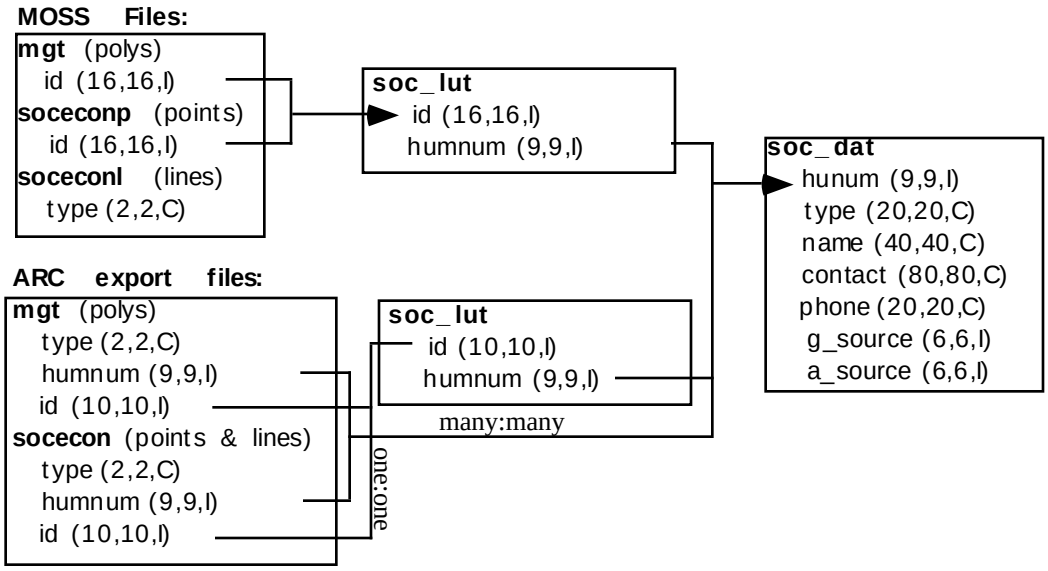


Figure 3: Links to the Human Use attributes (The ARC files refer to either the double precision source files or the single precision "desktop" files.)

Note in figure 3 that the ARC files include an item called type. This field contains a two character abbreviation describing the kind of resource that's referenced. See the *Environmental Sensitivity Index Guidelines* for a list of abbreviations.

As mentioned earlier, MOSS Files only allow the export of one attribute per file. In addition, MOSS files can contain only one feature type, i.e. points, lines or polygons. The *socecon* layer contains both points and arcs so each of these features is exported separately for the MOSS files (*soceconl and soceconp)* . The arc version is exported with the attribute type. The point version, as well as the *mgt* layer, are exported with the attribute id that links to the *soc_dat* table.

Base Map Layers

The base map layers, *hydro, esi* and *index* do not link to any additional data tables, but there are items directly associated with the layers. The ARC export files retain all of the attribute data, however the MOSS files require selection of the highest priority attribute for export, as again, MOSS supports only one attribute per layer. Shown in figure 4 are the attributes associated with each of the base layers. The ARC files show the layers *esi,* and *esil*. For ARC/INFO users, the cover *esi* will contain the arc attributes in the .aat table and the polygon attributes in the .pat table.

Attributes for both polygons and arcs may also be retained with some translators. However, with ArcView 3 and MOSS, it is possible to only query/display one feature type per layer. Therefore, in order to access the esi line attribute values as well as the polygon values, it is necessary to export a second cover with just an .aat file (esil). The polygon feature is given higher priority in ArcvView 3, so the polygon attributes will be what is displayed for *esi*. The arc and polygon features are exported seperately for the MOSS files as *esip* and *esil*. The *hydro* layer is only exported once as an ARC file, as users who are limited to displaying one feature type will have sufficient information with the polygonal value, water_code. For a description of the contents of the base layer fields, please see the *Environmental Sensitivity Index Guidelines*.

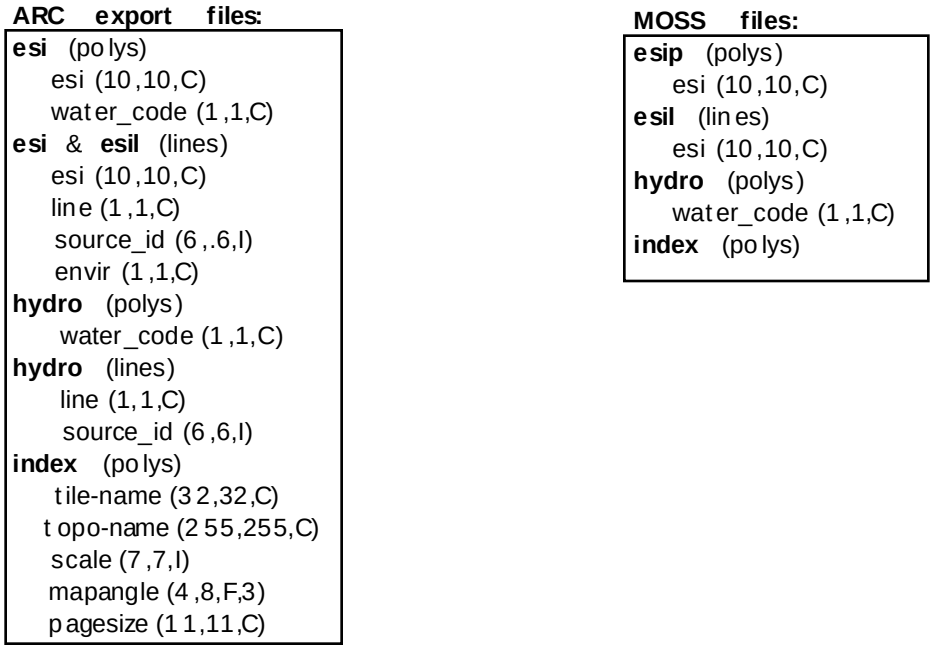


Figure 4: The base map layers and associated attributes

A modification from the description in the Guidelines was made to the esi polygon values. In the event an esi was undefined, the value for water_code was copied to the esi field. This means the water_code value in the MOSS files is basically retained since the water_code value is implicit or implied based on the esi value. Moving this value to esi is also an aid when trying to display maps with the proper colors and symbology. See the next section for further details.

Displaying ESI Maps

One of NOAA's goals, when creating the ESI standards discussed in the *Environmental Sensitivity Index Guidelines*, was that users would be able to create their own maps using standardized colors and symbolization making them universally understandable by other ESI users. Extending this capability to desktop GIS systems is becoming more practical and some basic approaches are outlined below. Standard colors, hatch patterns and the ESI symbolset, are outlined in the Guidelines in chapter 5.

Base Map layers

The *hydro* layer is color coded based on the value of water_code. Since the hard copy maps use the scanned quad sheets as a backdrop for the esi map, no standard color has been established for land or water. At this time it is suggested that the user select a subtle shade of green to greenish-brown for the land and a shade of blue for the water. The values for water_code are 'W', water, and 'L', land.

Both the *esi* polygons and the *esi* lines are shaded based on the value of the item esi. RGB and CMYK colors are listed in chapter 5 of the Guidelines for all possible values of esi, with two exceptions. As explained in the previous section, esi may also be assigned a value of 'L' or 'W'. For these polygons, shade them using the same color scheme as was selected for the *hydro* layer.

Display difficulties arise when an arc is assigned multiple esi values, i.e. 10/3. For the hard copy maps, the width of the line is divided into the number of different esi values present, and, starting from the landward side, that portion of the line is shaded with the corresponding esi color. These maps are produced using ARC/INFO and an ARC symbolset and lookup table are provided on the desktop CDs for ARC users. The solution in other mapping programs is less intuitive. ESRI expects that someday ArcView will be able to utilize ARC linesets and symbolsets which would solve the problem for those users. All users may want to contact the vendors of their particular software to see if they support custom line and symbolsets. Meanwhile, one "solution" is to code the the line with the color of the most sensitive esi value. The true value of the line could still be queried, but would not be immediately recognizable. If any users have other software specific suggestions or experience interest on the part of

some of their vendors, please contact Jill Petersen at j_petersen@hazmat.noaa.gov.

The *index* layer basically shows the boundaries of the USGS quads that were used when the hard copy maps were produced. Displaying the *index* simply as unfilled polygons may provide a locational reference to users. It also will help those who want to cross reference their digital data with the original atlas.

Biology layers

The color and hatch schemes for the various biological layers are also outlined in chapter 5 of the Guidelines. Basically the user will have one color and one hatch pattern per layer. On the hard copy maps, if polygons of multiple elements overlapped, the polygon would be shown as black. If a user wants to print a map, they may want to try and emulate this practice. However for those wishing to use the digital data for query etc., the overlapping colors will be of little consequence.

It's important to realize that while a version of these digital data were used to produce the hard copy ESI maps, it was not without significant manipulation to the digital data causing them to appear differently and, in some cases, resulting in loss of information. For example, a hard copy ESI map usually has at least one common throughout box. For each of the rarnums represented in that box, a polygon has been removed from the map itself. This results in less clutter and easier to read maps. However, if someone wanted to query a GIS system to find out the areal extent of endangered turtles, the answer wouldn't include the polygon that was moved to the common throughout designation. It helps to remember the analysis and query functions are a very significant part of desktop GIS.

However a user may want to make a customized map, for example showing the polygons where the endangered turtles may be found in February. Once the polygons are displayed in their GIS program, they may want to go in and edit the data to create a more pleasant appearing final product. This may involve creating common throughout boxes, dissolving adjacent polygons and/or adding symbolization. The ability to perform edits will be closely tied to the mapping software being used.

An ESI symbolset is provided on the desktop CD. For ARC users included is the entire symbol, line and shadeset files, as well as a lookup table that will help in color coding the ESI layer. A readme file is included for those users. For others, the symbolset only is provided as a true type font. Again, you may want to check with your software vendor to see if there is a way to integrate such a font into your mapping system. There is a subset of the ESI symbols available in ArcView 3. A print out of the symbolset is shown in chapter 5 of the Guidelines.

Human Use

The human use symbols should reflect the value listed in type. Reference appendix B of the Guidelines to see the valid values for socecon and mgt item type. These can then be assigned the appropriate symbols from the esi symbolset shown in chapter 5.

Reference

NOAA. 1997. Environmental Sensitivity Index Guidelines, Version 2. NOAA Technical Memorandum NOS ORCA 115. Seattle: Hazardous Materials Response and Assessment Division, National Oceanic and Atmospheric Administration.

MOSS STANDARD INPUT FILE FORMAT

FILE TYPE: STANDARD MOSS INPUT
FILE LENGTH: VARIABLE

RECORD TYPE: HEADER (REPEATING SERIES OF HEADER RECORDS PRE-
CEEDING ASSOCIATED COORDINATE PAIR DATA)
(READ IN I5,10X,15A2,5X,I5 FORMAT)
RECORD LENGTH: 56 CHARACTERS
RECORD POSITION: (1) BEFORE START OF ASSOCIATED COORDINATE DATA

*WORD NUMBER	BYTE NUMBER	FIELD LENGTH	FIELD TYPE	FIELD DESCRIPTION
1-3	1-5	5	CHAR	Item Number(NEGATIVE IF THE COORDINATES ARE LON/LAT)
3-8	6-15	10	CHAR	Blanks
8-23	16-45	30	CHAR	Attribute Name
23-25	46-50	5	CHAR	Blanks
26-28	51-55	5	CHAR	Number of coord. pairs
28	56	1	CHAR	ASCII NEW LINE CHARACTER

FILE TYPE: STANDARD MOSS INPUT
FILE LENGTH: VARIABLE LENGTH

RECORD TYPE: COORDINATE PAIRS (SERIES OF X,Y COORDINATES FOLLOWING HEADER RECORD)
RECORD LENGTH: 23 CHARACTERS
RECORD POSITION: AFTER HEADER RECORD (NUMBER OF COORDINATE RECORDS EQUALS THE VALUE INDICATED BY BYTES 51-55 IN THE HEADER RECORD) (READ IN 2F11.2 FORMAT)

*WORD NUMBER	BYTE NUMBER	FIELD LENGTH	FIELD TYPE	FIELD DESCRIPTION
1-6	1-11	11	CHAR	X Coordinate
6-11	12-22	11	CHAR	Y Coordinate
12	23	1	CHAR	ASCII NEW LINE CHARACTER

IF COORDINATES ARE LONGITUDE/LATITUDE (READ IN 2F10.5,I2 FORMAT)

1-5	1-10	10	CHAR	LONGITUDE
6-10	11-20	10	CHAR	LATITUDE
11	21-22	2	CHAR	FLAG 0-NORMAL 1-INDICATES FIRST POINT OF ISLAND POLYGON
12	23	1	CHAR	ASCII NEW LINE CHARACTER

*NOTE: Assumes 2 Alpha characters (2 bytes) to one word.

MARPLOT™

MARPLOT is a general-purpose mapping application program that runs on both Macintosh and Windows. It is designed to be easy to use, fast, and to consume as little disk and memory space as possible, allowing you to create, view, and modify maps quickly and easily. It also allows you to link objects on your computer maps to data in other programs.

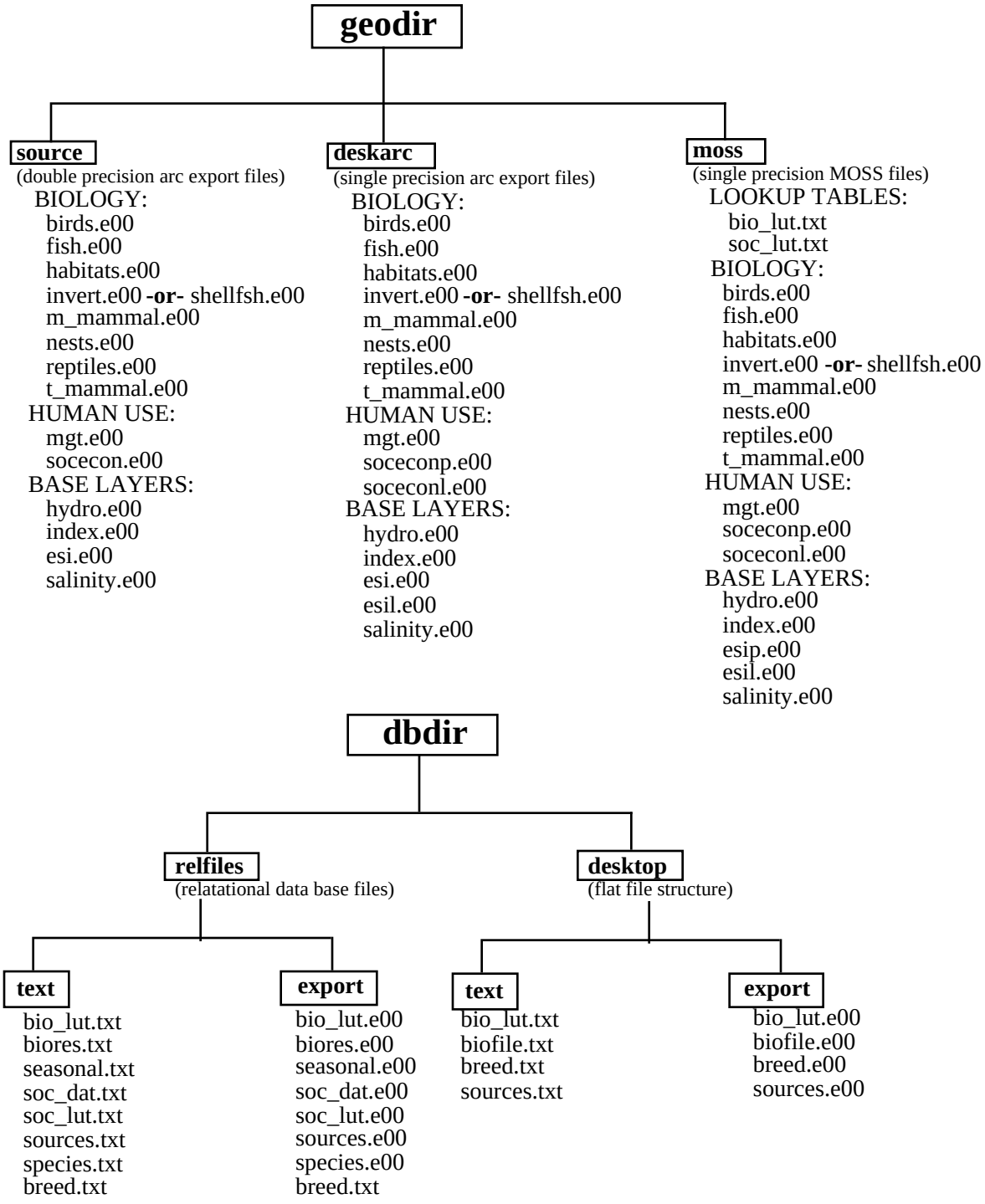
MAP DATA for MARPLOT comes from a variety of sources. All of the TIGER/Line® data from the Bureau of the Census (roads, water bodies, railroads, parks, etc.) is available in MARPLOT format on 10 LandView™ CDs. (These CD's also contain many EPA-regulated sites, demographic data , and several types of geographic boundaries including states, counties, cities, congressional districts, etc.) Other source data, in a number of formats, can be translated easily into MARPLOT files. The MARPLOT files themselves are compact, logically organized, and platform-independent.

SEVEN TYPES OF OBJECTS make up the content of MARPLOT maps. These are: points (symbols), rectangles, circles, polygons, polylines, text labels, and pictures. You might use a point object to mark the location of a building or monitoring site. Polylines are used to represent things like roads and streams. Polygons are used to represent things like parks or water bodies. Picture objects allow you to take any raster (bitmap, PICT) image and display it at fixed geographical coordinates. MARPLOT provides functions for creating, examining, and modifying each type of object.

MAPS AND LAYERS are the groups into which objects are organized. Layers usually contain a certain type of object, such as roads, water bodies, hospitals, etc. Maps, which usually cover a specific geographic area, contain some number of layers, with layers typically spanning multiple maps. There is no limit to the number of maps and layers that can be open and displayed simultaneously in MARPLOT. It is easy to put a given map/layer in or out of use at a particular MARPLOT installation, or to transfer a map/layer from one installation to another.

SEARCHING FOR MAP OBJECTS based on geographical criteria is easy and fast in MARPLOT. You can ask questions ranging in complexity from "What objects are at this point?" to "How many objects on one of these three layers are within 1.5 miles of this threat zone?"

INTER-APPLICATION COMMUNICATION allows MARPLOT to share information with databases and other programs. These programs contain information about map objects. Users can query the database for certain records, and then show the selected records on the map. From the map, users can select certain objects, and then get information from the database for those objects. Examples of database programs that communicate with MARPLOT are CAMEO and LandView. **CAMEO**, developed by NOAA and EPA, is a tools for planning for and responding to hazardous chemical emergencies. MARPLOT is used to plot the locations of facilities, hospitals, and other sites of interest to planners and responders, and to examine the geographical extents of real or imagined emergencies. **LandView**, developed by EPA, is a collection of data for EPA-regulated sites and demographic and economic information from the 1990 census, combined with a program for exploring the database and for displaying data in MARPLOT thematically according to your queries.

Directory Structure for ESI Desktop CD's

The master directory, named for the Atlas, will contain the geodir & dbdir directories